

Re: Shorter checksum than MD5

Source: <http://coding.derkeiler.com/Archive/Python/comp.lang.python/2004-09/1871.html>

From: Elbert Lev (elbertlev_at_hotmail.com)

Date: 09/10/04

Date: 10 Sep 2004 05:27:30 -0700

> *For your application, you should consider the total number of records
> you ever expect to have, and use more than $2 * \lg(\text{records})$ bits of hash.
> Due to the so-called "birthday paradox", when you have N possible hash
> values, two will be identical with 50% probability with around \sqrt{N}
> items. You'd probably prefer that the probability be much lower in your
> application, since a collision will result in incorrect results.
>*

Wrong! "birthday paradox" is not applicable here.

If you want an analogy with this combinatorial problem, imagine 2 rows with N objects in each, There exists a "measure" of each object.

Some objects can be modified with probability $1/2^{**32}$

the measure will not change after modification.

Objects in the SAME POSITION in each row are compared by comparing their measures.

After M objects are modified what is the probability that

at least one modification will be "missed" by the comparison process.

I don't think, that in the foreseen future (if M and N are not too high)

such collision will occur.